

# Phase Transition in Limiting Distributions of Coherence of High-Dimensional Random Matrices

Tony Cai<sup>1</sup> and Tiefeng Jiang<sup>2</sup>

University of Pennsylvania and University of Minnesota

## Abstract

The coherence of a random matrix, which is defined to be the largest magnitude of the Pearson correlation coefficients between the columns of the random matrix, is an important quantity for a wide range of applications including high-dimensional statistics and signal processing. Inspired by these applications, this paper studies the limiting laws of the coherence of  $n \times p$  random matrices for a full range of the dimension  $p$  with a special focus on the ultra high-dimensional setting. Assuming the columns of the random matrix are independent random vectors with a common spherical distribution, we give a complete characterization of the behavior of the limiting distributions of the coherence. More specifically, the limiting distributions of the coherence are derived separately for three regimes:  $\frac{1}{n} \log p \rightarrow 0$ ,  $\frac{1}{n} \log p \rightarrow \beta \in (0, \infty)$ , and  $\frac{1}{n} \log p \rightarrow \infty$ . The results show that the limiting behavior of the coherence differs significantly in different regimes and exhibits interesting phase transition phenomena as the dimension  $p$  grows as a function of  $n$ . Applications to statistics and compressed sensing in the ultra high-dimensional setting are also discussed.

**Keywords:** Coherence, correlation coefficient, limiting distribution, maximum, phase transition, random matrix, sample correlation matrix, Chen-Stein method.

**AMS 2000 Subject Classification:** Primary 62H12, 60F05; secondary 60F15, 62H10.

---

<sup>1</sup>Statistics Department, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, tcai@wharton.upenn.edu. The research of Tony Cai was supported in part by NSF FRG Grant DMS-0854973.

<sup>2</sup>School of Statistics, University of Minnesota, 224 Church Street, MN55455, tjiang@stat.umn.edu. The research of Tiefeng Jiang was supported in part by NSF FRG Grant DMS-0449365.

# 1 Introduction

With dramatic advances in computing and technology, large and high-dimensional datasets are now routinely collected in many scientific investigations. The associated statistical inference problems, where the dimension  $p$  can be much larger than the sample size  $n$ , arise naturally in a wide range of applications including compressed sensing, climate studies, genomics, functional magnetic resonance imaging, risk management and portfolio allocation. Conventional statistical methods and results based on fixed  $p$  and large  $n$  are no longer applicable and these applications call for new technical tools and new statistical procedures.

The coherence of a random matrix, which is defined to be the largest magnitude of the off-diagonal entries of the sample correlation matrix generated from the random matrix, has been shown to be an important quantity for many applications. For example, the coherence has been used for testing the covariance structure of high-dimensional distributions (Cai and Jiang (2010)), the construction of compressed sensing matrices and high dimensional regression in statistics (see, e.g., Candes and Tao (2005), Donoho, Elad and Temlyakov (2006) and Cai, Wang and Xu (2010a, b)). In addition, the coherence has also been used in signal processing, medical imaging, and seismology. Some of these problems are seemingly unrelated at first sight, but interestingly they can all be attacked through the use of the limiting laws of the coherence of random matrices (see, e.g., Cai and Jiang (2010)). In these applications, a case of special interest is when the dimension  $p$  is much larger than the sample size  $n$ . Indeed, in compressed sensing and other related problems the goal is often to make the dimension  $p$  as large as possible relative to the sample size  $n$ .

In the present paper we study the limiting laws of the coherence of random matrices. Let  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$  and  $\mathbf{y} = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ . Recall the Pearson correlation coefficient  $\rho$  defined by

$$\rho = \rho_{\mathbf{x}, \mathbf{y}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  and  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ . Let  $\mathbf{X}_1, \dots, \mathbf{X}_p$  be independent  $n$ -dimensional random vectors, and let  $\rho_{ij}$  be the correlation coefficient between  $\mathbf{X}_i$  and  $\mathbf{X}_j$ . Set  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p) = (x_{ij})_{n \times p}$ . The coherence of the random matrix  $\mathbf{X}$  is defined as

$$L_n = \max_{1 \leq i < j \leq p} |\rho_{ij}|. \quad (2)$$

In certain applications such as the construction of compressed sensing matrices, the means  $\mu_i = E\mathbf{X}_i$  and  $\mu_j = E\mathbf{X}_j$  are given and one is interested in

$$\tilde{\rho}_{ij} = \frac{(\mathbf{X}_i - \mu_i)^T (\mathbf{X}_j - \mu_j)}{\|\mathbf{X}_i - \mu_i\| \cdot \|\mathbf{X}_j - \mu_j\|}, \quad 1 \leq i, j \leq p \quad (3)$$

and the corresponding coherence is defined by

$$\tilde{L}_n = \max_{1 \leq i < j \leq p} |\tilde{\rho}_{ij}|. \quad (4)$$

The goal of this paper is to give a complete characterization of the behavior of the limiting distributions of  $L_n$  and  $\tilde{L}_n$  over the full range of  $p$  (as a function of  $n$ ) including the super-exponential case where  $(\log p)/n \rightarrow \infty$ .

The coherence  $L_n$  has been studied intensively in recent years. Jiang (2004) was the first to show that if  $x_{ij}$ 's are independent and identically distributed (i.i.d.) with  $E|x_{ij}|^{30+\epsilon} < \infty$  for some  $\epsilon > 0$  and  $n/p \rightarrow \gamma \in (0, \infty)$ , then  $nL_n^2 - 4 \log p + \log \log p$  converges weakly to an extreme distribution of type I with distribution function

$$F(y) = e^{-\frac{1}{\sqrt{8\pi}}e^{-y/2}}, \quad y \in \mathbb{R}. \quad (5)$$

Throughout this paper,  $\log x = \log_e x$  for any  $x > 0$  and  $p = p_n$  depends on  $n$  only. The result (5) was later improved in several papers by sharpening the moment assumptions and relaxing the restrictions between  $n$  and  $p$ . In terms of the relationship between  $n$  and  $p$ , these results can be classified into the following categories:

- (a). *Linear rate:*  $p \sim cn$  with  $c$  being a constant. Li and Rosalsky (2006), Zhou (2007), Li, Liu and Rosalsky (2009) and Li, Qi and Rosalsky (2010) improved the moment conditions to make (5) valid under the condition  $p/n \rightarrow c \in (0, 1)$ .
- (b). *Polynomial rate:*  $p = O(n^\alpha)$  with  $\alpha > 0$  being a constant. Liu, Lin and Shao (2008) showed that (5) holds as  $p \rightarrow \infty$  and  $p = O(n^\alpha)$  where  $\alpha$  is a constant. That is, (5) still holds when  $n$  and  $p$  are in the polynomial rates.
- (c). *Sub-exponential rate:*  $\log p = o(n^\alpha)$  with  $0 < \alpha \leq 1/3$  being a constant. Motivated by applications in testing high-dimensional covariance structure and construction of compressed sensing, Cai and Jiang (2010) further extended the range of  $p$  by considering the sub-exponential rate. It was shown that (5) is also valid if  $\log p = o(n^\alpha)$  with  $\alpha \in (0, 1/3]$  and the distribution of  $x_{11}$  is well-behaved. In particular, (5) holds with  $\alpha = 1/3$  when  $x_{ij}$ 's are i.i.d.  $N(0, 1)$  random variables.

An interesting question is whether the limiting distribution (5) holds for even higher dimensional case when  $\log p$  is of order  $n^\alpha$  with  $\alpha > 1/3$ . This is a case of significant interest in high-dimensional data analysis and signal processing. For example, in the context of high-dimensional regression and classification, simulation studies about the distribution of  $L_n$  were made in Cai and Lv (2007) and Fan and Lv (2008 and 2010). In this paper we shall study the limiting laws of the coherence  $L_n$  for a full range of the values of  $p$ . To make our technical analysis tractable, we focus on the setting where the columns  $\mathbf{X}_i$  of the random matrix  $\mathbf{X}$  follow a spherical distribution, which contains the normal distribution  $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$

as a special case. Motivated by the applications in statistics and signal processing mentioned earlier, we are especially interested in the ultra high dimensional case. More specifically, we consider three different regimes:

- (i). *the sub-exponential case*:  $\frac{1}{n} \log p \rightarrow 0$ ;
- (ii). *the exponential case*:  $\frac{1}{n} \log p \rightarrow \beta \in (0, \infty)$ ;
- (iii). *the super-exponential case*:  $\frac{1}{n} \log p \rightarrow \infty$ .

Our results show that the limiting behavior of  $L_n$  differs significantly in different regimes and exhibits interesting phase transition phenomena as the dimension  $p$  grows as a function of  $n$ . To answer the question posed earlier, it is shown that  $nL_n^2 - 4 \log p + \log \log p$  converges to the limiting distribution given in (5) if and only if  $\log p = o(n^{1/2})$ . The phase transition in the limiting distribution first occurs with the case when  $\log p$  is of order  $n^{1/2}$ . In this transitional case, additional shift in the limiting distribution occurs. When the dimension  $p$  further grows as a function of  $n$ , another transition occurs in the range when  $\log p$  is of the same order as  $n$ . In the sub-exponential case,  $L_n$  converges to 0 in probability. When  $\log p \sim \beta n$  for some positive constant  $\beta$ ,  $L_n$  converges in probability to a constant strictly between 0 and 1, and the limiting distribution of  $T_n = \log(1 - L_n^2)$  is significantly different from that in the sub-exponential case. If  $p$  is further increased to the super-exponential case,  $L_n$  converges to 1 in probability and the limiting distribution of  $T_n$  becomes the extreme value distribution without a shift.

There are also interesting differences between the limiting behaviors of  $L_n$  and  $\tilde{L}_n$ . As shown in Cai and Jiang (2010), the limiting laws of  $L_n$  and  $\tilde{L}_n$  coincide with each other when  $x_{ij}$ 's are iid  $N(0, 1)$  variables and  $\log p = o(n^{1/3})$ . Our results show that this remains true in the current setting for the sub-exponential and exponential cases, but not true for the super-exponential case. It is interesting to contrast the results obtained in this paper with the results on  $L_n$  and  $\tilde{L}_n$  in the previous literature. The only known limiting distribution of  $L_n$  and  $\tilde{L}_n$  is given in (5) and the best known result in terms of the range of  $p$  is  $\log p = o(n^{1/3})$ . In comparison, our study significantly extends the knowledge on the limiting distributions of the coherence and shows the “colorful” phase transition phenomena as the dimension  $p$  increases.

The limiting laws of the coherence have immediate applications in statistics and signal processing. Testing the covariance structure of a high dimensional random variable is an important problem in statistical inference. A particularly interesting problem is to test for independence in the Gaussian case because many statistical procedures are built upon the assumptions of independence and normality of the observations. The limiting laws of the coherence derived in this paper can be used directly to construct a test for independence in the ultra high dimensional setting. In addition, the limiting laws can also be used for the construction of compressed sensing matrices. We shall discuss these applications in Section 3.

Many sophisticated probabilistic tools have been used in the previous literature to study the limiting laws of the coherence. For example, the Chen-Stein method, large deviation inequalities, and strong approximations were used to derive the results mentioned earlier in (a), (b) and (c). Yet there appears to be limitations to these methods. It is unclear (to us) whether these techniques can be easily adopted to derive the limiting distribution of  $L_n$  when  $\log p$  is of order  $n^\alpha$  for  $\alpha > 1/3$  and answer the question posed earlier. See Remark 4.1 in Section 4 for further discussions. In this paper a different technique is developed. Under the assumption that  $\mathbf{X}_i$  in (2) has a spherical distribution, we first show a somewhat surprising result that the sample correlation coefficients  $\{\rho_{ij}; 1 \leq i < j \leq p\}$  are pairwise independent. We then apply the Chen-Stein method to the coherence  $L_n = \max_{1 \leq i < j \leq p} |\rho_{ij}|$  by using the exact distribution of  $\rho_{ij}$  and the pairwise-independence structure of  $\rho_{ij}$ . In addition, the exact distribution of  $\rho_{ij}$  also leads to some interesting properties of  $\rho_{ij}$  in the small sample cases:  $\rho_{ij}$  has the *symmetric Bernoulli distribution* for  $n = 2$ , that is,  $P(\rho_{ij} = \pm 1) = 1/2$ ;  $\rho_{ij}^2$  follows the *Arcsine law* on  $[0, 1]$  for  $n = 3$ ;  $\rho_{ij}$  follows the *uniform distribution* on  $[-1, 1]$  for  $n = 4$ ; and  $\rho_{ij}$  follows the *semi-circle law* for  $n = 5$ .

The rest of the paper is organized as follows. Section 2 studies the limiting laws of the coherence  $L_n$  and  $\tilde{L}_n$  of a random matrix in the high-dimensional setting under the three regimes. The interesting phase transition phenomena are discussed in detail. Section 3 considers two direct applications of the limiting laws derived in this paper to statistics and signal processing in the ultra high dimensional setting. Section 4 discusses some of the interesting aspects of the techniques used in the derivations. Connections and differences with other related work, for example, the relationship between the sample correlation coefficients and the angles between random vectors, are discussed in Section 5. The main results are proved in Section 6.

## 2 Limiting Laws of the Coherence

In this section we study separately the limiting behaviors of the coherence  $L_n$  and  $\tilde{L}_n$  of an  $n \times p$  random matrix  $\mathbf{X}$  under the three regimes:  $\frac{1}{n} \log p \rightarrow 0$ ,  $\frac{1}{n} \log p \rightarrow \beta \in (0, \infty)$ , and  $\frac{1}{n} \log p \rightarrow \infty$ . As mentioned before, we shall focus on the setting where the columns  $\mathbf{X}_i$  of the random matrix  $\mathbf{X}$  follow a spherical distribution.

### 2.1 Limiting Laws of the Coherence $L_n$

A random vector  $\mathbf{Y} \in \mathbb{R}^n$  is said to follow a *spherical distribution* if  $\mathbf{OY}$  and  $\mathbf{Y}$  have the same probability distribution for all  $n \times n$  orthogonal matrix  $\mathbf{O}$ . Examples of spherical distributions include:

- the multivariate normal distribution  $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$  with  $\sigma > 0$ ;

- the normal scale-mixutre distribution  $\sum_{k=1}^K \epsilon_k N(\mathbf{0}, \sigma_k^2 \mathbf{I}_n)$  with the density function

$$\sum_{k=1}^K \epsilon_k \frac{1}{(2\pi\sigma_k^2)^{n/2}} \cdot \exp\left(-\frac{1}{2\sigma_k^2} \mathbf{y}^T \mathbf{y}\right) \quad (6)$$

where  $\sigma_k > 0$ ,  $\epsilon_k > 0$ , and  $\sum_{k=1}^K \epsilon_k = 1$ ;

- the multivariate  $t$  distribution with  $m$  degrees of freedom and density function

$$\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})(m\pi)^{n/2}} \cdot \left(1 + \frac{1}{m} \mathbf{y}^T \mathbf{y}\right)^{(m+n)/2} \quad (7)$$

for  $m \geq 1$ . The case  $m = 1$  corresponds to the multivariate Cauchy distribution.

See Muirhead (1982) for further discussions on spherical distributions.

Let  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p) = (x_{ij})_{n \times p}$  be an  $n \times p$  random matrix. Throughout the rest of this paper, we shall assume:

**Assumption (A):** the columns  $\mathbf{X}_1, \dots, \mathbf{X}_p$  are independent  $n$ -dimensional random vectors with a common spherical distribution (which may depend on  $n$ ) and  $P(\mathbf{X}_1 = \mathbf{0}) = 0$ .

The condition  $P(\mathbf{X}_1 = \mathbf{0}) = 0$  is to ensure that the correlation coefficients are well defined. Let  $\rho_{ij}$  be the Pearson correlation coefficient of  $\mathbf{X}_i$  and  $\mathbf{X}_j$  for  $1 \leq i < j \leq p$ . Then,  $\Psi_n := (\rho_{ij})_{p \times p}$  is the correlation matrix of  $\mathbf{X}$ , and  $L_n$  defined in (2), is the largest magnitude of the off-diagonal entries of the sample correlation matrix  $\Psi_n$ .

To make the statements of the limiting distributions uniform across different regimes, we shall state all the results in the main theorems in terms of  $T_n = \log(1 - L_n^2)$ . We begin with the sub-exponential case.

**THEOREM 1 (Sub-Exponential Case)** *Suppose  $p = p_n$  satisfies  $(\log p)/n \rightarrow 0$  as  $n \rightarrow \infty$ , then under Assumption (A),*

- (i).  $L_n \rightarrow 0$  in probability as  $n \rightarrow \infty$ .
- (ii). Let  $T_n = \log(1 - L_n^2)$ . Then, as  $n \rightarrow \infty$ ,

$$nT_n + 4 \log p - \log \log p \quad (8)$$

*converges weakly to an extreme distribution with the distribution function  $F(y) = 1 - e^{-Ke^{y/2}}$ ,  $y \in \mathbb{R}$  and  $K = 1/\sqrt{8\pi}$ .*

The following law of large numbers is a direct consequence of Theorem 1.

**COROLLARY 2.1** *Assume the same conditions as in Theorem 1, we have*

$$\sqrt{\frac{n}{\log p}} L_n \rightarrow 2 \quad (9)$$

*in probability as  $n \rightarrow \infty$ .*

This result actually provides the convergence speed of  $L_n \rightarrow 0$  stated in Theorem 1(i). It is stronger than Theorem 2 in Cai and Jiang (2010), which shows (9) holds if  $\log p = o(n^{1/3})$  and  $x_{ij}$ 's are i.i.d.  $N(0, 1)$  random variables.

Theorem 1 also shows an interesting phase transition phenomenon of the limiting behavior of the coherence  $L_n$ .

**COROLLARY 2.2 (Transitional Case)** *Suppose  $p = p_n$  satisfies  $\lim_{n \rightarrow \infty} (\log p)/\sqrt{n} = \alpha \in [0, \infty)$ , then under Assumption (A),*

$$nL_n^2 - 4 \log p + \log \log p \quad (10)$$

*converges weakly to the distribution function  $\exp\{-\frac{1}{\sqrt{8\pi}}e^{-(y+8\alpha^2)/2}\}$ ,  $y \in \mathbb{R}$ .*

As mentioned in the introduction, Cai and Jiang (2010) shows that  $nL_n^2 - 4 \log p + \log \log p$  converges weakly to an extreme distribution with distribution function given in (5) when  $\log p = o(n^{1/3})$  and  $x_{ij}$  are independent standard normal variables. This is the best known result in the literature in terms of the range of  $p$ . Corollary 2.2 shows that (5) holds if and only if  $\log p = o(n^{1/2})$  when  $\mathbf{X}_1$  has a spherical distribution which includes the normal distribution  $N(0, I_n)$  as a special case. This answers the question asked earlier in this paper. Corollary 2.2 also shows that the limiting distribution of  $L_n$  has a transitional phase between  $(\log p)/\sqrt{n} \rightarrow 0$  and  $(\log p)/\sqrt{n} \rightarrow \infty$ . In the transitional case when  $(\log p)/\sqrt{n} \rightarrow \alpha \in (0, \infty)$ , the limiting distribution of  $nL_n^2 - 4 \log p + \log \log p$  is shifted to the left by  $8\alpha^2$ .

We now consider the exponential case.

**THEOREM 2 (Exponential Case)** *Suppose  $p = p_n$  satisfies  $(\log p)/n \rightarrow \beta \in (0, \infty)$  as  $n \rightarrow \infty$ , then under Assumption (A),*

(i).  $L_n \rightarrow \sqrt{1 - e^{-4\beta}}$  in probability as  $n \rightarrow \infty$ .

(ii). Let  $T_n = \log(1 - L_n^2)$ . Then, as  $n \rightarrow \infty$ ,

$$nT_n + 4 \log p - \log \log p \quad (11)$$

*converges weakly to the distribution function*

$$F(y) = 1 - \exp\left\{-K(\beta)e^{(y+8\beta)/2}\right\}, \quad y \in \mathbb{R}, \quad \text{where } K(\beta) = \left(\frac{\beta}{2\pi(1 - e^{-4\beta})}\right)^{1/2}. \quad (12)$$

Theorem 2 reveals the behavior of  $L_n$  in the transitional case  $(\log p)/n \rightarrow \beta$ . In this case, the coherence  $L_n$  converges in probability to a constant strictly between 0 and 1. Dividing (11) by  $n$ , it is easy to see that

$$T_n \rightarrow -4\beta \text{ in probability as } n \rightarrow \infty$$

since  $\lim_{n \rightarrow \infty} (\log p)/n = \beta \in (0, \infty)$ . This is also a direct consequence of Theorem 2(i). Furthermore, it is trivially true that  $1 - e^{-4\beta} \sim 4\beta$  as  $\beta \rightarrow 0^+$ . Thus,

$$\lim_{\beta \rightarrow 0^+} K(\beta) = \frac{1}{\sqrt{8\pi}},$$

which is exactly the value of  $K$  in Theorem 1. Thus, the limiting distribution  $F(y)$  in Theorem 2 as  $\beta \rightarrow 0^+$  becomes the limiting distribution  $F(y)$  in Theorem 1. Heuristically, the sub-exponential case covered in Theorem 1 corresponds to the case “ $\beta = 0$ ” in Theorem 2. On the other hand, the exponential case of  $(\log p)/n \rightarrow \beta \in (0, \infty)$  can also be viewed as a transitional phase between the sub-exponential and super-exponential cases.

Finally we turn to the super-exponential case where  $(\log p)/n \rightarrow \infty$ .

**THEOREM 3 (Super-Exponential Case)** *Suppose  $p = p_n$  satisfies  $(\log p)/n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $T_n = \log(1 - L_n^2)$ . Then under Assumption (A),*

- (i).  $L_n \rightarrow 1$  in probability as  $n \rightarrow \infty$ . Further,  $\frac{n}{\log p} T_n \rightarrow -4$  in probability as  $n \rightarrow \infty$ .
- (ii). As  $n \rightarrow \infty$ ,

$$nT_n + \frac{4n}{n-2} \log p - \log n \tag{13}$$

converges weakly to the distribution function  $F(y) = 1 - e^{-Ke^{y/2}}$ ,  $y \in \mathbb{R}$  with  $K = 1/\sqrt{2\pi}$ .

The correction term of  $nT_n$  in (13) is  $\frac{4n}{n-2} \log p - \log n$ , which is different from the term “ $4 \log p - \log \log p$ ” appeared in (8) and (11). A reason is that  $T_n$  converges to a finite constant in probability in Theorems 1 and 2, whereas  $T_n$  goes to  $-\infty$  in probability in Theorem 3. On the other hand, suppose  $(\log p)/n \rightarrow \beta \in (0, \infty)$  and  $\beta$  is large, then  $\log n = \log \log p - \log \beta + o(1)$  and

$$\frac{4n}{n-2} \log p = 4 \log p + \frac{8}{n-2} \log p = 4 \log p + 8\beta + o(1)$$

as  $n \rightarrow \infty$ . Consequently, the quantity in (13) becomes

$$(nT_n + 4 \log p - \log \log p) + \text{constant} + o(1)$$

as  $n \rightarrow \infty$ . The part in the parenthesis is the same as (8) in Theorem 1 and (11) in Theorem 2. This says that, heuristically, the results in Theorems 1, 2 and 3 are consistent.

The formulation in the above theorems is in terms of  $T_n = \log(1 - L_n^2)$  for uniformity. However, one can easily change the expressions in terms of the coherence  $L_n$ . For instance,

$$P(n \log(1 - L_n^2) + 4 \log p - \log \log p \leq y) = P(L_n \geq \sqrt{s_n})$$

where

$$s_n := 1 - \exp \left\{ \frac{1}{n} (-4 \log p + \log \log p + y) \right\}. \tag{14}$$



## 2.2 Limiting Laws of $\tilde{L}_n$

We now study the limiting laws of the coherence  $\tilde{L}_n$  defined in (3) and (4). Note that under Assumption (A), the columns  $\mathbf{X}_1, \dots, \mathbf{X}_p$  are independent  $n$ -dimensional random vectors with a common spherical distribution. By symmetry, it is easy to see that the mean  $\mu = E\mathbf{X}_i = \mathbf{0}$  if it exists and hence

$$\tilde{\rho}_{ij} = \frac{\mathbf{X}_i^T \mathbf{X}_j}{\|\mathbf{X}_i\| \cdot \|\mathbf{X}_j\|} \quad \text{and} \quad \tilde{L}_n = \max_{1 \leq i < j \leq p} |\tilde{\rho}_{ij}|. \quad (15)$$

As mentioned in the introduction, Cai and Jiang (2010) showed that the limiting laws of  $L_n$  and  $\tilde{L}_n$  coincide with each other when  $x_{ij}$ 's are iid  $N(0, 1)$  random variables and  $\log p = o(n^{1/3})$ . We shall show that this is still true in our current setting for the sub-exponential and exponential cases, but not true for the super-exponential case.

**THEOREM 4 (Sub-Exponential & Exponential Cases)** *Under the same conditions, Theorems 1 and 2 and Corollaries 2.1 and 2.2 hold with  $L_n$  replaced by  $\tilde{L}_n$ .*

In the super-exponential case, the limiting behaviors of  $\tilde{L}_n$  and  $L_n$  are different.

**THEOREM 5 (Super-Exponential Case)** *Suppose  $p = p_n$  satisfies  $(\log p)/n \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $\tilde{T}_n = \log(1 - \tilde{L}_n^2)$ . Then under Assumption (A),*

- (i).  $\tilde{L}_n \rightarrow 1$  in probability as  $n \rightarrow \infty$ . Further,  $\frac{n}{\log p} \tilde{T}_n \rightarrow -4$  in probability as  $n \rightarrow \infty$ .
- (ii). As  $n \rightarrow \infty$ ,

$$n\tilde{T}_n + \frac{4n}{n-1} \log p - \log n \quad (16)$$

converges weakly to the distribution function  $F(y) = 1 - e^{-Ke^{y/2}}$ ,  $y \in \mathbb{R}$  with  $K = 1/\sqrt{2\pi}$ .

Note the difference between (13) and (16). When  $(\log p)/n \rightarrow \infty$ , the difference between  $\frac{4n}{n-2} \log p$  and  $\frac{4n}{n-1} \log p$  is not negligible.

## 3 Applications

As mentioned in the introduction, the limiting laws of the coherence have a wide range of applications. Here we discuss briefly two immediate applications, one in high-dimensional statistics and another in signal processing. These applications were also discussed in Cai and Jiang (2010), but restricted to the Gaussian case with  $\log p = o(n^{1/3})$ . Here we extend to the more general spherical distributions and higher dimensions.

Testing the covariance structure of a distribution is an important problem in high dimensional statistical inference. Let  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  be a random sample from a  $p$ -variate spherical

distribution with covariance matrix  $\Sigma_{p \times p} = (\sigma_{ij})$ . We wish to test the hypotheses that  $\Sigma$  is diagonal, i.e.,

$$H_0 : \sigma_{i,j} = 0 \text{ for all } |i - j| \geq 1 \text{ vs. } H_a : \sigma_{i,j} \neq 0 \text{ for some } |i - j| \geq 1. \quad (17)$$

In the Gaussian case, this is the same as testing for independence. The asymptotic distribution of  $L_n$  can be used to construct a convenient test statistic for testing the hypotheses in (17). For example, in the case  $\log p = o(n^{1/2})$ , an approximate level  $\alpha$  test is to reject the null hypothesis  $H_0$  whenever

$$L_n^2 \geq n^{-1} \left( 4 \log p - \log \log p - \log(8\pi) - 2 \log \log(1 - \alpha)^{-1} \right).$$

It follows directly from Theorem 1 that the size of this test goes to  $\alpha$  asymptotically as  $n \rightarrow \infty$ . This test was introduced in Cai and Jiang (2010) in the Gaussian case with the restriction that  $\log p = o(n^{1/3})$ .

Similarly, in the exponential (and sub-exponential) case, set

$$D_{n,p} = nT_n + 4 \log p - \log \log p.$$

Then Theorem 2 states that

$$P(D_{n,p} \leq y) \rightarrow 1 - \exp \left( -K(\beta) e^{(y+8\beta)/2} \right), \quad (18)$$

where  $K(\beta) = \left( \frac{\beta}{2\pi(1-e^{-4\beta})} \right)^{1/2}$ . An approximate level  $\alpha$  test for testing the hypotheses in (17) can be obtained by rejecting the null hypothesis  $H_0$  whenever

$$D_{n,p} \leq 2 \log \log(1 - \alpha)^{-1} - 2 \log K(\beta) - 8\beta.$$

A test for the super-exponential case can also be constructed analogously by using the limiting distribution given in Theorem 3.

Compressed sensing is an active and fast growing field in signal processing. See, e.g., Donoho (2006), Candes and Tao (2007), Bickel, Ritov and Tsybakov (2009), Candes and Plan (2009), and Cai, Wang and Xu (2010a, b). An important problem in compressed sensing is the construction of measurement matrices  $\mathbf{X}_{n \times p}$  which enables the precise recovery of a sparse signal  $\beta$  from linear measurements  $\mathbf{y} = \mathbf{X}\beta$  using an efficient recovery algorithm. Such a measurement matrix  $\mathbf{X}$  is typically randomly generated because it is difficult to construct deterministically. The best known example is perhaps the  $n \times p$  random matrix  $\mathbf{X}$  whose entries  $x_{i,j}$  are iid normal variables

$$x_{i,j} \stackrel{iid}{\sim} N(0, n^{-1}). \quad (19)$$

A commonly used condition is the mutual incoherence property (MIP) which requires the pairwise correlations among the column vectors of  $\mathbf{X}$  to be small. Write  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)$  =

$(x_{ij})_{n \times p}$  with  $x_{ij}$  satisfying (19) and let the coherence  $\tilde{L}_n = \max_{1 \leq i < j \leq p} |\tilde{\rho}_{ij}|$  be defined as in (3) and (4). It has been shown that the condition

$$(2k - 1)\tilde{L}_n < 1 \quad (20)$$

ensures the exact recovery of  $k$ -sparse signal  $\beta$  in the noiseless case where  $y = X\beta$  (see Donoho and Huo (2001) and Fuchs (2004)), and stable recovery of sparse signal in the noisy case where

$$y = X\beta + z.$$

Here  $z$  is an error vector, not necessarily random. See Cai, Wang and Xu (2010b).

The limiting laws derived in this paper can be used to show how likely a random matrix satisfies the MIP condition (20). Take the sub-exponential case as an example. By Theorem 4, as long as  $(\log p)/n \rightarrow 0$ ,

$$\tilde{L}_n \sim 2\sqrt{\frac{\log p}{n}}.$$

So in order for the MIP condition (20) to hold, roughly the sparsity  $k$  should satisfy

$$k < \frac{1}{4}\sqrt{\frac{n}{\log p}}.$$

## 4 Technical Tool: Distribution of Correlation Coefficients

In this section we shall discuss the methodology used in our technical arguments. Sophisticated approximation methods such as the Chen-Stein method, large deviation bounds and strong approximations are the main ingredients in the proofs of the previous results in the literature including those given in Jiang (2004), Li and Rosalsky (2006), Zhou (2007), Liu, Lin and Shao (2008), Li, Liu and Rosalsky (2009), Li, Qi and Rosalsky (2010), and Cai and Jiang (2010). Though these technical tools work well for the cases when the dimension  $p$  is not ultra high, it is far from clear to us whether/how these same tools can be used to derive the limiting distributions of the coherence  $L_n$  for the three regimes considered in Section 2.

In this paper, a different approach is developed to derive the limiting distributions of  $L_n$ . Assuming the  $\mathbf{X}_i$ 's have the spherical distribution, we find an interesting and useful property of the correlation coefficients  $\{\rho_{ij}; 1 \leq i < j \leq p\}$  and  $\{\tilde{\rho}_{ij}; 1 \leq i < j \leq p\}$  given below.

**LEMMA 4.1** *Let  $n \geq 3$ . Under Assumption (A), the Pearson correlation coefficient  $\{\rho_{ij}; 1 \leq i < j \leq p\}$  are pairwise independent and identically distributed with density function*

$$f(\rho) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \cdot (1 - \rho^2)^{\frac{n-4}{2}}, \quad |\rho| < 1. \quad (21)$$

Similarly,  $\{\tilde{\rho}_{ij}; 1 \leq i < j \leq p\}$  are pairwise independent and identically distributed with density

$$g(\rho) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \cdot (1 - \rho^2)^{\frac{n-3}{2}}, \quad |\rho| < 1. \quad (22)$$

Note that the only difference between (21) and (22) is the “degree of freedom”: replacing  $n$  in (22) with  $n-1$ , one gets (21). This is not difficult to understand by noting the definition of  $\rho_{ij} = \frac{(\mathbf{X}_i - \bar{\mathbf{X}})^T (\mathbf{X}_j - \bar{\mathbf{X}})}{\|\mathbf{X}_i - \bar{\mathbf{X}}\| \cdot \|\mathbf{X}_j - \bar{\mathbf{X}}\|}$ . Heuristically, by subtracting  $\bar{\mathbf{X}}$  from  $\mathbf{X}_i$ , the distribution of  $\rho_{ij}$  becomes one degree less than that of  $\tilde{\rho}_{ij} = \frac{\mathbf{X}_i^T \mathbf{X}_j}{\|\mathbf{X}_i\| \cdot \|\mathbf{X}_j\|}$ .

Although  $\{\rho_{ij}; 1 \leq i < j \leq p\}$  are pairwise independent, they are not mutually independent. In fact, recalling  $\Psi = \Psi_n = (\rho_{ij})_{p \times p}$ , the probability density function of  $\Psi$  is given by

$$h(\Psi) = B_{n,p} \cdot (\det(\Psi))^{(n-p-2)/2} \quad (|\rho_{ij}| < 1, i < j) \quad (23)$$

for  $1 \leq p < n$ , where  $B_{n,p}$  is an (explicit) normalizing constant, see p.148 from Muirhead (1982). Obviously,  $h(\Psi)$  is not a product of functions of individual  $\rho_{ij}$ ’s, the entries of  $\Psi$ , hence  $\{\rho_{ij}; 1 \leq i < j \leq p\}$  are not independent.

Lemma 4.1 also yields the following interesting results on the distribution of the correlation coefficients  $\rho_{ij}$  in the small sample cases. The verification is given in Section 6.

**COROLLARY 4.1** *Under Assumption (A), the following holds for all  $1 \leq i < j \leq p$ .*

- (i). *When  $n = 2$ ,  $\rho_{ij}$  has the symmetric Bernoulli distribution, i.e.,  $P(\rho_{ij} = \pm 1) = 1/2$ .*
- (ii). *When  $n = 3$ ,  $\rho_{ij}$  has the density  $f(\rho) = \frac{1}{\pi} \frac{1}{\sqrt{1-\rho^2}}$  on  $(-1, 1)$ . That is,  $\rho_{ij}^2$  follows the arcsine law on  $[0, 1]$ .*
- (iii). *When  $n = 4$ ,  $\rho_{ij}$  follows the uniform distribution on  $[-1, 1]$ .*
- (iv). *When  $n = 5$ ,  $\rho_{ij}$  has the density  $f(\rho) = \frac{2}{\pi} \sqrt{1-\rho^2}$  for  $|\rho| \leq 1$ . That is,  $\rho_{ij}$  follows the semi-circle law.*

Lemma 4.1 provides a major technical tool for the proof of the main results. The starting step in the proofs of our theorems is the Chen-Stein method (Lemma 6.3) which requires the evaluation of two quantities:  $P(\rho_{ij} \geq C)$  and  $P(\rho_{ij} \geq C, \rho_{kl} \geq C)$ . By using the explicit density expression in (21), we are able to evaluate the first probability precisely. The pairwise independence stated in Lemma 4.1 yields  $P(\rho_{ij} \geq C, \rho_{kl} \geq C) = P(\rho_{ij} \geq C)^2$  for  $\{i, j\} \neq \{k, l\}$ . In other words, the evaluation of the second quantity is reduced to the study of the first one. This greatly simplifies some of the technical arguments.

**REMARK 4.1** Equation (21) yields directly that  $W_n := \sqrt{n}\rho_{12}$  has the density function

$$f_n(w) = \frac{1}{\sqrt{n}} \cdot \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \cdot \left(1 - \frac{w^2}{n}\right)^{\frac{n-4}{2}} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-w^2/2}$$

as  $n \rightarrow \infty$  for all  $w \in \mathbb{R}$ , where the fact that  $\Gamma(\frac{n-1}{2})/\Gamma(\frac{n-2}{2}) \sim \sqrt{n/2}$  as  $n \rightarrow \infty$  (see (33)) is used. This shows that  $W_n$  converges to  $N(0, 1)$  in distribution as  $n \rightarrow \infty$ . Set  $(x_{ij})_{n \times p} := (\mathbf{X}_1, \dots, \mathbf{X}_p)$ . Assuming that  $x_{ij}$ 's are i.i.d. with an unknown distribution but with suitable moment conditions, say,  $|x_{12}|$  is bounded, it can be shown easily that  $\sqrt{n}\rho_{12}$  converges to  $N(0, 1)$  by using the standard central limit theorem for i.i.d. random variables and the Slutsky theorem. However, the convergence speed is hard to be captured well enough so that  $L_n$  in (2) is understood clearly when  $p$  is much larger than  $n$ . The best known result is that (5) holds for  $\log p = o(n^\alpha)$  with  $\alpha = 1/3$  in Cai and Jiang (2010). Here, with the understanding of the pairwise independence among  $\{\rho_{ij}; 1 \leq i < j \leq p\}$  and the exact distribution of  $\rho_{ij}$  we are able to get the limiting distribution of  $L_n$  for the full range of the values of  $p$  and to fully characterize the phase transition phenomena in the limiting behaviors of the coherence (Theorems 1, 2 and 3 and the corresponding corollaries).

## 5 Discussions

The present paper was inspired by the applications in high-dimensional statistics and signal processing in which the dimension  $p$  is often desired to be as high as possible as a function of  $n$ . All the known results on the coherence  $L_n$  are restricted to the cases where the dimension  $p$  is either linear, polynomial or at most sub-exponential in  $n$ . In comparison, we give in this paper a complete characterization of the limiting distribution of  $L_n$  for the full range of  $p$  including the sub-exponential case  $\frac{1}{n} \log p \rightarrow 0$ , the exponential case  $\frac{1}{n} \log p \rightarrow \beta \in (0, \infty)$ , and the super-exponential case  $\frac{1}{n} \log p \rightarrow \infty$ . Our results show interesting phase transition phenomena in the limiting distributions of the coherence when the dimension  $p$  grows as a function of  $n$ . Over the full range of values of  $p$ , phase transition of the limiting behavior of  $L_n$  occurs twice: when  $\log p$  is of order  $n^{1/2}$  and when  $\log p$  is of order  $n$ . These results also show that the standard limiting distribution (5) known in the literature holds if and only if  $\log p = o(n^{1/2})$  when the columns have a spherical distribution which includes the commonly considered i.i.d. normal setting as a special case.

Previous results on the coherence  $L_n$  focus on the case where the entries  $x_{ij}$  of the random matrix  $\mathbf{X}$  are i.i.d. under certain moment conditions. See the references mentioned in (a), (b) and (c) in the introduction. In this paper, we assume the columns of  $\mathbf{X} = (x_{ij})_{n \times p}$  to be i.i.d. with a spherical distribution. The spherical distribution assumption are more special than the non-specified distributions with certain moment conditions considered in the previous literature. On the other hand, the entries of a vector with a spherical distribution do not have to be independent (see, e.g., the normal scale-mixture distribution in (6)

and the multivariate  $t$ -distribution in (7)). In this sense, our work relaxes the independence assumption among the entries  $x_{ij}$ . Under the assumption of spherical distributions, we are able to show that the sample correlation coefficients are pairwise independent and then use the exact distribution and the pairwise-independence structure of the sample correlation coefficients as a major technical tool in the derivation of the limiting distributions.

There are interesting connections between sample correlation coefficients and angles between random vectors. Let  $\mathbf{a} \in \mathbb{R}^n$  be a deterministic vector with  $\|\mathbf{a}\| = 1$ . Let  $\mathbf{X}_1 \in \mathbb{R}^n$  be a random vector with a spherical distribution satisfying  $P(\mathbf{X}_1 = \mathbf{0}) = 0$ . Relating Theorem 1.5.7(i) and (5) on page 147 in Muirhead (1982), it can be seen that  $W = \frac{\mathbf{a}^T \mathbf{X}_1}{\|\mathbf{X}_1\|}$  has the same distribution as the one given in (22). Note that  $\frac{\mathbf{X}_1}{\|\mathbf{X}_1\|}$  has the uniform distribution over the unit sphere in  $\mathbb{R}^n$ , and hence  $W$  is the cosine of the angle between a fixed unit vector  $\mathbf{a}$  and a random vector with the uniform distribution on the unit sphere. Similar to Corollary 4.1, the following holds.

- (i). If  $n = 2$ , then the cosine of the angle has the probability density function  $f(\rho) = \frac{1}{\pi} \frac{1}{\sqrt{1-\rho^2}}$ . That is, the square of the cosine follows the Arcsine law on  $[0, 1]$ .
- (ii). If  $n = 3$ , then the cosine of the angle follows the uniform distribution on  $[-1, 1]$ .
- (iii). If  $n = 4$ , then the cosine of the angle has the probability density function  $f(\rho) = \frac{2}{\pi} \sqrt{1-\rho^2}$  for  $|\rho| \leq 1$ . That is,  $\rho_{ij}$  follows the semi-circle law.

The semi-circle law is perhaps best known in random matrix theory as the limit of the empirical distribution of the eigenvalues of an  $n \times n$  Wigner random matrix as  $n \rightarrow \infty$ . See, e.g., Wigner (1958). It seems not so common to see a random variable to satisfy the semi-circle law in practice. It is interesting to see the semi-circle law here as the exact distribution of the correlation coefficient and the cosine of the angle between two random vectors in Corollary 4.1(iv) and (iii) above.

## 6 Proofs

In this section we prove the main results of the paper. We shall write  $p$  for  $p_n$  if there is no confusion. We begin by proving Lemma 4.1 on the distributions of the correlation coefficients. We then collect and prove a few additional technical results before giving the proofs of the main theorems.

### 6.1 Technical Results

The following lemma is needed for the proof of Lemma 4.1.

**LEMMA 6.1** Let  $\mathbf{X}$  be an  $n$ -dimensional random vector with a spherical distribution and  $P(\mathbf{X} = \mathbf{0}) = 0$ . Let  $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$  and  $\{1\} = \{k\mathbf{1}; k \in \mathbb{R}\}$ , the span of  $\mathbf{1}$ . Then  $P(\mathbf{X} \in \{1\}) = 0$ .

**Proof.** Since  $P(\mathbf{X} = \mathbf{0}) = 0$ , we know  $\mathbf{Y} := \frac{\mathbf{X}}{\|\mathbf{X}\|}$  is well-defined. By definition,  $\mathbf{O}\mathbf{X} \stackrel{P}{=} \mathbf{X}$  for any orthogonal matrix  $\mathbf{O}$ , then

$$\mathbf{O}\mathbf{Y} = \frac{\mathbf{O}\mathbf{X}}{\|\mathbf{O}\mathbf{X}\|} \stackrel{P}{=} \frac{\mathbf{X}}{\|\mathbf{X}\|} = \mathbf{Y}.$$

That is, the probability measure generated by  $\mathbf{Y}$  is an orthogonal-invariant measure on the unit sphere  $S^{n-1} \subset \mathbb{R}^n$ . Since the Haar probability measure, as the distribution on the unit sphere with the orthogonal-invariant property, is unique, it follows that  $\mathbf{Y}$  must have the uniform distribution on the unit sphere in  $\mathbb{R}^n$ . In particular,  $P(\mathbf{Y} = y) = 0$  for any  $y \in S^{n-1}$ . Let  $A = \{\mathbf{X} \in \{1\} \setminus \{\mathbf{0}\}\}$  and  $y_0 = n^{-1/2}(1, \dots, 1)^T \in S^{n-1}$ . Notice  $A \subset \{\mathbf{Y} = y_0 \text{ or } -y_0\}$ . It follows that  $P(\mathbf{X} \in \{1\}) = P(A) \leq P(\mathbf{Y} = y_0) + P(\mathbf{Y} = -y_0) = 0$ . ■

**Proof of Lemma 4.1.** Recall that  $\mathbf{X}_1, \dots, \mathbf{X}_p$  are independent and  $\rho_{ij}$  is the Pearson correlation coefficient of  $\mathbf{X}_i$  and  $\mathbf{X}_j$  for  $1 \leq i < j \leq p$ . Given  $i < j$  and  $k < l$  with  $(i, j) \neq (k, l)$ . It is easy to see that  $\rho_{ij}$  and  $\rho_{kl}$  are independent if  $\{i, j\} \cap \{k, l\} = \emptyset$ . Thus, to finish the proof, it enough to prove the following:

Let  $\{\mathbf{U}, \mathbf{V}, \mathbf{W}\}$  be i.i.d with an  $n$ -dimensional spherical distribution and  $P(\mathbf{U} = \mathbf{0}) = 0$ .

Then  $\rho_{\mathbf{U}, \mathbf{V}}$  and  $\rho_{\mathbf{U}, \mathbf{W}}$  are i.i.d. with the density function given in (21). (24)

By Lemma 6.1,  $P(\mathbf{U} \in \{1\}) = P(\mathbf{V} \in \{1\}) = P(\mathbf{W} \in \{1\}) = 0$ . Then,  $\rho_{\mathbf{U}, \mathbf{V}}$  and  $\rho_{\mathbf{U}, \mathbf{W}}$  have the same probability density function  $f(\rho)$  by (5) on p. 147 from Muirhead (1982). To show the independence, we need to prove

$$E[g(\rho_{\mathbf{U}, \mathbf{V}}) \cdot h(\rho_{\mathbf{U}, \mathbf{W}})] = E g(\rho_{\mathbf{U}, \mathbf{V}}) \cdot E h(\rho_{\mathbf{U}, \mathbf{W}}) \quad (25)$$

for any bounded and measurable functions  $g(x)$  and  $h(x)$ . Since  $\mathbf{U}$ ,  $\mathbf{V}$  and  $\mathbf{W}$  are independent,

$$\begin{aligned} E[g(\rho_{\mathbf{U}, \mathbf{V}}) \cdot h(\rho_{\mathbf{U}, \mathbf{W}})] &= E\left\{E[g(\rho_{\mathbf{U}, \mathbf{V}}) \cdot h(\rho_{\mathbf{U}, \mathbf{W}})|\mathbf{U}]\right\} \\ &= E\left\{E[g(\rho_{\mathbf{U}, \mathbf{V}})|\mathbf{U}] \cdot E[h(\rho_{\mathbf{U}, \mathbf{W}})|\mathbf{U}]\right\}. \end{aligned} \quad (26)$$

Write  $\mathbf{V} = (V_1, \dots, V_n)^T \in \mathbb{R}^n$  and  $\bar{\mathbf{V}} = \frac{1}{n} \sum_{i=1}^n V_i$ . For any numbers  $u_1, \dots, u_n$  such that at least two of them are not identical, Theorem 5.1.1 and (5) on p. 147 from Muirhead (1982) say that

$$\rho_{\mathbf{u}, \mathbf{V}} = \frac{\sum_{i=1}^n (u_i - \bar{\mathbf{u}})(V_i - \bar{\mathbf{V}})}{\sqrt{\sum_{i=1}^n (u_i - \bar{\mathbf{u}})^2 \cdot \sum_{i=1}^n (V_i - \bar{\mathbf{V}})^2}}$$

has the probability density function  $f(\rho)$  as in (21), where  $\mathbf{u} = (u_1, \dots, u_n)^T$  and  $\bar{\mathbf{u}} = \frac{1}{n} \sum_{i=1}^n u_i$  (see also Kariya and Eaton (1977) for this). In other words, given  $\mathbf{U}$ , the probability distribution of  $\rho_{\mathbf{U}, \mathbf{V}}$  does not depend on the value of  $\mathbf{U}$ . Let  $\mathbf{U} = (U_1, \dots, U_n)^T$ . Evidently,  $P(U_1 = \dots = U_n) = P(\mathbf{U} \in \{1\}) = 0$ . Thus,

$$E[g(\rho_{\mathbf{U}, \mathbf{V}})|\mathbf{U}] = \int_{|\rho| \leq 1} g(\rho) f(\rho) d\rho = Eg(\rho_{\mathbf{U}, \mathbf{V}})$$

and

$$E[h(\rho_{\mathbf{U}, \mathbf{W}})|\mathbf{U}] = Eh(\rho_{\mathbf{U}, \mathbf{W}})$$

since  $\rho_{\mathbf{U}, \mathbf{V}}$  and  $\rho_{\mathbf{U}, \mathbf{W}}$  have the same probability density function  $f(\rho)$  as in (21). These and (26) conclude (25).

We now turn to study  $\tilde{\rho}_{ij}$ . Given  $1 \leq i < j \leq p$ . Then  $\alpha := \frac{\mathbf{X}_j}{\|\mathbf{X}_j\|}$  is a unit vector and is independent of  $\mathbf{X}_i$ . Further,  $\tilde{\rho}_{ij} = \frac{\alpha^T \mathbf{X}_i}{\|\mathbf{X}_i\|}$ . It then follows from Theorem 1.5.7(i) and the argument for (5) on p.147 of Muirhead (1982) that  $\tilde{\rho}_{ij}$  has the probability density function  $f(\rho)$  as in (22). The proof for the pairwise independence among  $\{\tilde{\rho}_{ij}; 1 \leq i < j \leq p\}$  is the same as that for the  $\rho_{ij}$ 's. ■

**Proof of Corollary 4.1.** Taking  $n = 3, 4, 5$ , respectively, in Lemma 4.1, we easily have (ii), (iii) and (iv). Now we check (i).

Let  $\mathbf{X}_1 = (\xi_1, \eta_1)^T \in \mathbb{R}^2$  and  $\mathbf{X}_2 = (\xi_2, \eta_2)^T \in \mathbb{R}^2$ . It is easy to see

$$\rho_{12} = \frac{\xi_1 - \eta_1}{|\xi_1 - \eta_1|} \cdot \frac{\xi_2 - \eta_2}{|\xi_2 - \eta_2|}. \quad (27)$$

First, Assumption (A) and Lemma 6.1 imply  $P(\xi_i = \eta_i) = 0$  for  $i = 1, 2$ . Since  $\mathbf{X}_1$  has a spherical distribution, we know that  $\mathbf{A}\mathbf{X}_1$  and  $\mathbf{X}_1$  have the same distribution for any  $\mathbf{A} = \text{diag}(\epsilon_1, \epsilon_2)$  with  $\epsilon_i = \pm 1$ ,  $i = 1, 2$ . This implies  $\mathbf{X}_1$  is symmetric, and hence  $\xi_1 - \eta_1$  is symmetric. Consequently,  $\frac{\xi_1 - \eta_1}{|\xi_1 - \eta_1|}$  takes value  $\pm 1$  with probability  $1/2$  each. The same is true for  $\frac{\xi_2 - \eta_2}{|\xi_2 - \eta_2|}$ . By (27) and the independence between  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , we conclude that  $P(\rho_{12} = \pm 1) = 1/2$ . ■

**LEMMA 6.2** *Let  $t = t_m \in (0, 1)$  satisfy  $mt_m^2 \rightarrow \infty$  as  $m \rightarrow \infty$ . Then*

$$\int_t^1 (1 - x^2)^{m/2} dx = \frac{1}{mt} (1 - t^2)^{(m+2)/2} (1 + o(1))$$

as  $m \rightarrow \infty$ .

**Proof.** Set  $y = x^2$  for  $x > 0$ . Then  $x = \sqrt{y}$  and

$$\begin{aligned} I_m := \int_t^1 (1 - x^2)^{m/2} dx &= \frac{1}{2} \int_{t^2}^1 \frac{1}{\sqrt{y}} (1 - y)^{m/2} dy \\ &= -\frac{1}{m+2} \int_{t^2}^1 \frac{1}{\sqrt{y}} [(1 - y)^{(m+2)/2}]' dy. \end{aligned} \quad (28)$$



By integration by parts,

$$\begin{aligned} I_m &= -\frac{1}{m+2} \frac{1}{\sqrt{y}} (1-y)^{(m+2)/2} \Big|_{t^2}^1 - \frac{1}{2(m+2)} \int_{t^2}^1 \frac{1}{y^{3/2}} (1-y)^{(m+2)/2} dy \\ &= \frac{1}{(m+2)t} (1-t^2)^{(m+2)/2} - \frac{1}{m+2} \cdot \frac{1}{2} \int_{t^2}^1 \frac{1}{\sqrt{y}} (1-y)^{m/2} \cdot \frac{1-y}{y} dy. \end{aligned} \quad (29)$$

Note that  $0 \leq \frac{1-y}{y} \leq \frac{1}{t^2}$  for all  $[t^2, 1]$ . By the second equality in (28),

$$0 < \frac{1}{m+2} \cdot \frac{1}{2} \int_{t^2}^1 \frac{1}{\sqrt{y}} (1-y)^{m/2} \cdot \frac{1-y}{y} dy \leq \frac{1}{mt^2} I_m.$$

This and (29) conclude that

$$\frac{1}{(m+2)t} (1-t^2)^{(m+2)/2} - \frac{1}{mt^2} I_m \leq I_m \leq \frac{1}{(m+2)t} (1-t^2)^{(m+2)/2}.$$

Solving the first inequality on  $I_m$ , we have

$$\left(1 + \frac{1}{mt^2}\right)^{-1} \frac{1}{(m+2)t} (1-t^2)^{(m+2)/2} \leq I_m \leq \frac{1}{(m+2)t} (1-t^2)^{(m+2)/2}.$$

By the given condition that  $mt^2 = mt_m^2 \rightarrow \infty$ , we arrive at

$$I_m = \frac{1}{(m+2)t} (1-t^2)^{(m+2)/2} (1 + o(1)) = \frac{1}{mt} (1-t^2)^{(m+2)/2} (1 + o(1))$$

as  $m \rightarrow \infty$ . ■

The following Poisson approximation result is essentially a special case of Theorem 1 from Arratia et al. (1989).

**LEMMA 6.3** *Let  $I$  be an index set and  $\{B_\alpha, \alpha \in I\}$  be a set of subsets of  $I$ , that is,  $B_\alpha \subset I$  for each  $\alpha \in I$ . Let also  $\{\eta_\alpha, \alpha \in I\}$  be random variables. For a given  $t \in \mathbb{R}$ , set  $\lambda = \sum_{\alpha \in I} P(\eta_\alpha > t)$ . Then*

$$|P(\max_{\alpha \in I} \eta_\alpha \leq t) - e^{-\lambda}| \leq (1 \wedge \lambda^{-1})(b_1 + b_2 + b_3)$$

where

$$\begin{aligned} b_1 &= \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} P(\eta_\alpha > t) P(\eta_\beta > t), \quad b_2 = \sum_{\alpha \in I} \sum_{\alpha \neq \beta \in B_\alpha} P(\eta_\alpha > t, \eta_\beta > t), \\ b_3 &= \sum_{\alpha \in I} E|P(\eta_\alpha > t | \sigma(\eta_\beta, \beta \notin B_\alpha)) - P(\eta_\alpha > t)|, \end{aligned}$$

and  $\sigma(\eta_\beta, \beta \notin B_\alpha)$  is the  $\sigma$ -algebra generated by  $\{\eta_\beta, \beta \notin B_\alpha\}$ . In particular, if  $\eta_\alpha$  is independent of  $\{\eta_\beta, \beta \notin B_\alpha\}$  for each  $\alpha$ , then  $b_3 = 0$ .

**LEMMA 6.4** *Let  $L_n$  be as in (2) and Assumption (A) hold. For  $\{t_n \in [0, 1]; n \geq 1\}$ , set*

$$h_n = \frac{n^{1/2}p^2}{\sqrt{2\pi}} \int_{t_n}^1 (1-x^2)^{\frac{n-4}{2}} dx, \quad n \geq 1.$$

*If  $\lim_{n \rightarrow \infty} h_n = \lambda \in [0, \infty)$ , then  $\lim_{n \rightarrow \infty} P(L_n \leq t_n) = e^{-\lambda}$ .*

**Proof.** For brevity of notation, we sometimes write  $t = t_n$  if there is no confusion. First, take  $I = \{(i, j); 1 \leq i < j \leq p\}$ . For  $u = (i, j) \in I$ , set  $B_u = \{(k, l) \in I; \text{one of } k \text{ and } l = i \text{ or } j, \text{ but } (k, l) \neq u\}$ ,  $\eta_u = |\rho_{ij}|$  and  $A_u = A_{ij} = \{|\rho_{ij}| > t\}$ . By the i.i.d. assumption on  $\mathbf{X}_1, \dots, \mathbf{X}_p$  and Lemma 6.3,

$$|P(L_n \leq t) - e^{-\lambda_n}| \leq b_{1,n} + b_{2,n} \quad (30)$$

where

$$\lambda_n = \frac{p(p-1)}{2} P(A_{12}) \quad (31)$$

and

$$b_{1,n} \leq 2p^3 P(A_{12})^2 \text{ and } b_{2,n} \leq 2p^3 P(A_{12}A_{13}).$$

By Lemma 4.1,  $A_{12}$  and  $A_{13}$  are independent events with the same probability. Thus, from (31),

$$b_{1,n} \wedge b_{2,n} \leq 2p^3 P(A_{12})^2 \leq \frac{8p\lambda_n^2}{(p-1)^2} \leq \frac{32\lambda_n^2}{p} \quad (32)$$

for all  $p \geq 2$ . Now we compute  $P(A_{12})$ . In fact, by Lemma 4.1 again,

$$\begin{aligned} P(A_{12}) &= \int_{1 > |x| > t} f(x) dx = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \int_{1 > |x| > t} (1-x^2)^{\frac{n-4}{2}} dx \\ &= \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \int_t^1 (1-x^2)^{\frac{n-4}{2}} dx. \end{aligned}$$

Recalling the Stirling formula (see, e.g., p.368 from Gamelin (2001) or (37) on p.204 from Ahlfors (1979)):

$$\log \Gamma(z) = z \log z - z - \frac{1}{2} \log z + \log \sqrt{2\pi} + O\left(\frac{1}{x}\right)$$

as  $x = \operatorname{Re}(z) \rightarrow \infty$ , it is easy to verify that

$$\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-2}{2})} \sim \sqrt{\frac{n}{2}} \quad (33)$$

as  $n \rightarrow \infty$ . Thus,

$$P(A_{12}) \sim \frac{2n^{1/2}}{\sqrt{2\pi}} \int_t^1 (1-x^2)^{\frac{n-4}{2}} dx$$

as  $n \rightarrow \infty$ . From (31), we know

$$\lambda_n \sim \frac{n^{1/2}p^2}{\sqrt{2\pi}} \int_t^1 (1-x^2)^{\frac{n-4}{2}} dx = h_n$$

as  $n \rightarrow \infty$ . Finally, by (30) and (32), we know

$$\lim_{n \rightarrow \infty} P(L_n \leq t) = e^{-\lambda} \text{ if } \lim_{n \rightarrow \infty} h_n = \lambda \in [0, \infty). \quad \blacksquare$$

## 6.2 Proofs for Results on $L_n$ in Section 2.1

**Proof of Theorem 1.** (i). Assume (ii) of the theorem holds. Since  $(\log p)/n \rightarrow 0$  as  $n \rightarrow \infty$ , dividing (8) by  $n$ , we see that  $\log(1 - L_n^2) \rightarrow 0$  in probability, or equivalently,  $L_n \rightarrow 0$  in probability as  $n \rightarrow \infty$ .

(ii). The proof here does not rely on the conclusion in (i). We claim that

$$(n-2) \log(1 - L_n^2) + 4 \log p - \log \log p \quad (34)$$

converges weakly to the distribution function  $F(y) = 1 - e^{-Ke^{y/2}}$ ,  $y \in \mathbb{R}$ . Once this holds, using the condition that  $\log p = o(n)$  and the same argument as in (i) above, we have  $\log(1 - L_n^2) \rightarrow 0$  in probability as  $n \rightarrow \infty$ . Then by the Slutsky lemma,

$$n \log(1 - L_n^2) + 4 \log p - \log \log p$$

converges weakly to the distribution function  $F(y) = 1 - e^{-Ke^{y/2}}$ ,  $y \in \mathbb{R}$ . We then obtain (8). Now we prove (34).

Fix  $y \in \mathbb{R}$ . Let  $N = n - 2$  and  $t = t_n \in [0, 1)$  such that

$$\log(1 - t^2) = \frac{-4 \log p + \log \log p + y}{N} \wedge 0. \quad (35)$$

From (35) and the assumption  $\log p = o(n)$ , we have that  $t_n \rightarrow 0^+$  as  $n \rightarrow \infty$ , and hence  $\log(1 - t^2) \sim -t^2$ . Thus, (35) implies

$$t \sim \left( \frac{4 \log p - \log \log p - y}{N} \right)^{1/2} \sim \frac{2\sqrt{\log p}}{\sqrt{N}} \text{ and } Nt_n^2 \rightarrow \infty \quad (36)$$

as  $n \rightarrow \infty$ . By (35) again,

$$P((n-2) \log(1 - L_n^2) + 4 \log p - \log \log p \geq y) = P(L_n \leq t) \quad (37)$$

as  $n$  is large enough. Now let's compute  $h_n$  in Lemma 6.4 for  $\lim_{n \rightarrow \infty} P(L_n \leq t)$ . Recall

$$h_n = \frac{n^{1/2} p^2}{\sqrt{2\pi}} \int_t^1 (1-x^2)^{\frac{n-4}{2}} dx. \quad (38)$$

From Lemma 6.2 and the second assertion in (36),

$$\begin{aligned} n^{1/2} p^2 \int_t^1 (1-x^2)^{(n-4)/2} dx &\sim \frac{n^{1/2} p^2}{nt} (1-t^2)^{(n-2)/2} \\ &\sim \frac{p^2}{\sqrt{N}t} (1-t^2)^{N/2} \end{aligned}$$

as  $n \rightarrow \infty$ . This joint with (35) and the first assertion in (36) gives

$$\frac{p^2}{\sqrt{N}t} (1-t^2)^{N/2} \sim \frac{p^2}{2\sqrt{\log p}} \cdot \exp \left\{ \frac{-4 \log p + \log \log p + y}{N} \cdot \frac{N}{2} \right\} = \frac{1}{2} e^{y/2}$$

as  $n \rightarrow \infty$ . Combining the above three identities, we see that

$$h_n \rightarrow \frac{1}{\sqrt{8\pi}} e^{y/2}$$

as  $n \rightarrow \infty$ . Therefore, we conclude from Lemma 6.4 and (37) that

$$\lim_{n \rightarrow \infty} P((n-2) \log(1-L_n^2) + 4 \log p - \log \log p \geq y) = e^{-K e^{y/2}}$$

for any  $y \in \mathbb{R}$ , where  $K = \frac{1}{\sqrt{8\pi}}$ . Since  $\varphi(y) := e^{-K e^{y/2}}$  is continuous for all  $y \in \mathbb{R}$ , it is trivial to check that

$$\lim_{n \rightarrow \infty} P((n-2) \log(1-L_n^2) + 4 \log p - \log \log p \leq y) = 1 - e^{-K e^{y/2}} \quad (39)$$

for any  $y \in \mathbb{R}$ . We get (34).  $\blacksquare$

**Proof of Corollary 2.1.** Dividing (8) by  $\log p$ , we see that

$$\frac{n}{\log p} \log(1-L_n^2) \rightarrow -4 \quad (40)$$

in probability as  $n \rightarrow \infty$ . By (i) of Theorem 1, we know  $L_n \rightarrow 0$  in probability as  $n \rightarrow \infty$ . Since  $\rho_{ij}$  has density  $f(\rho)$  as in (21) for  $i \neq j$ , we have  $P(L_n = 0) = 0$  for all  $n \geq 3$ . Notice the function

$$h(x) := \begin{cases} x^{-1} \log(1-x), & \text{if } x \in (0, 1); \\ -1, & \text{if } x = 0 \end{cases}$$

is continuous on  $[0, 1)$ , we have

$$\frac{\log(1-L_n^2)}{L_n^2} = h(L_n^2) \rightarrow h(0) = -1$$

in probability as  $n \rightarrow \infty$ . This together with (40) yields

$$\frac{n}{\log p} \cdot L_n^2 \rightarrow 4$$

in probability as  $n \rightarrow \infty$ . The desired conclusion then follows.  $\blacksquare$

**Proof of Corollary 2.2.** By Theorem 1,

$$P(n \log(1 - L_n^2) + 4 \log p - \log \log p \leq y) \rightarrow F(y) \quad (41)$$

as  $n \rightarrow \infty$ , where  $F(y) = 1 - e^{-K e^{y/2}}$ ,  $y \in \mathbb{R}$ . Set

$$y_{n,p} = n \left[ 1 - \exp \left\{ \frac{1}{n} (-4 \log p + \log \log p + y) \right\} \right]. \quad (42)$$

Then, (41) becomes that  $P(n L_n^2 \geq y_{n,p}) \rightarrow F(y)$ , and hence

$$P(n L_n^2 - 4 \log p + \log \log p < y_{n,p} - 4 \log p + \log \log p) \rightarrow 1 - F(y) \quad (43)$$

as  $n \rightarrow \infty$  for any  $y \in \mathbb{R}$ . We claim

$$y_{n,p} - 4 \log p + \log \log p \rightarrow -(y + 8\alpha^2) \quad \text{if} \quad \frac{\log p}{\sqrt{n}} \rightarrow \alpha \in [0, \infty). \quad (44)$$

If this is true, by (43) and the continuity of  $F(y)$ ,

$$\lim_{n \rightarrow \infty} P(n L_n^2 - 4 \log p + \log \log p \leq -(y + 8\alpha^2)) = 1 - F(y)$$

for any  $y \in \mathbb{R}$ . In other words,  $n L_n^2 - 4 \log p + \log \log p$  converges weakly to a probability distribution function

$$G(z) := 1 - F(-z - 8\alpha^2) = \exp\{-K e^{-(z+8\alpha^2)/2}\}, \quad z \in \mathbb{R},$$

as  $n \rightarrow \infty$ . Now we prove claim (44).

In fact, set  $t = -4 \log p + \log \log p + y$ . Then  $t = O(\log p)$  and  $\frac{t}{n} \rightarrow 0$  as  $n \rightarrow \infty$  under the assumption  $\frac{\log p}{\sqrt{n}} \rightarrow \alpha$ . Consequently, by (42) and the Taylor expansion,

$$\begin{aligned} y_{n,p} = n(1 - e^{t/n}) &= -n \left[ \frac{t}{n} + \frac{t^2}{2n^2} + O\left(\frac{t^3}{n^3}\right) \right] \\ &= -t - \frac{t^2}{2n} + O\left(\frac{t^3}{n^2}\right) \end{aligned}$$

as  $n \rightarrow \infty$ . If  $\frac{\log p}{\sqrt{n}} \rightarrow \alpha$  as  $n \rightarrow \infty$ , then  $\frac{t^2}{2n} \rightarrow 8\alpha^2$  and  $\frac{t^3}{n^2} \rightarrow 0$  as  $n \rightarrow \infty$ . Therefore, (44) is concluded.  $\blacksquare$

**Proof of Theorem 2.** (i). Assume (ii) of the theorem holds. Since  $(\log p)/n \rightarrow \beta$  as  $n \rightarrow \infty$ , dividing (11) by  $n$ , we see that  $\log(1 - L_n^2) \rightarrow -4\beta$  in probability, or equivalently,  $L_n \rightarrow \sqrt{1 - e^{-4\beta}}$  in probability as  $n \rightarrow \infty$ .

(ii). The proof here does not rely on the conclusion in (i). We first show that

$$(n - 2) \log(1 - L_n^2) + 4 \log p - \log \log p \quad (45)$$

converges weakly to the distribution function  $F(y) = 1 - e^{-K(\beta)e^{y/2}}$ ,  $y \in \mathbb{R}$ , where  $K(\beta)$  is as in (12). If this is true, by the condition  $(\log p)/n \rightarrow \beta$  and the argument as in (i) above, we see that

$$\log(1 - L_n^2) \rightarrow -4\beta$$

in probability as  $n \rightarrow \infty$ . Thus, by the Slutsky lemma,

$$\begin{aligned} & n \log(1 - L_n^2) + 4 \log p - \log \log p \\ &= [(n - 2) \log(1 - L_n^2) + 4 \log p - \log \log p] + 2 \log(1 - L_n^2) \end{aligned}$$

converges weakly to the distribution function  $F(y) = 1 - e^{-K(\beta)e^{(y+8\beta)/2}}$ ,  $y \in \mathbb{R}$ . We now prove (45).

Fix  $y \in \mathbb{R}$ . Let  $N = n - 2$  and  $t = t_n \in [0, 1)$  such that

$$t^2 = 1 - \exp \left\{ \frac{1}{N} (-4 \log p + \log \log p + y) \wedge 0 \right\}.$$

It is easy to see that

$$P((n - 2) \log(1 - L_n^2) + 4 \log p - \log \log p \geq y) = P(L_n \leq t) \quad (46)$$

as  $n$  is sufficiently large, and

$$\lim_{n \rightarrow \infty} t_n = \sqrt{1 - e^{-4\beta}} \in (0, 1) \quad \text{and} \quad N \log(1 - t^2) = -4 \log p + \log \log p + y \quad (47)$$

as  $n$  is sufficiently large. We now calculate  $h_n$  in Lemma 6.4 to obtain  $\lim_{n \rightarrow \infty} P(L_n \leq t)$ .  
Review

$$h_n = \frac{n^{1/2} p^2}{\sqrt{2\pi}} \int_t^1 (1 - x^2)^{\frac{n-4}{2}} dx. \quad (48)$$

It follows from Lemma 6.2 and the first identity in (47) that

$$\begin{aligned} n^{1/2} p^2 \int_t^1 (1 - x^2)^{(n-4)/2} dx &\sim \frac{n^{1/2} p^2}{nt} (1 - t^2)^{(n-2)/2} \\ &\sim \frac{1}{\sqrt{1 - e^{-4\beta}}} \cdot \frac{p^2}{\sqrt{N}} (1 - t^2)^{N/2} \end{aligned}$$

as  $n \rightarrow \infty$ . By using the second identity in (47), we see that

$$\begin{aligned} \frac{p^2}{\sqrt{N}} (1 - t^2)^{N/2} &= \frac{p^2}{\sqrt{N}} \cdot \exp \left\{ \frac{-4 \log p + \log \log p + y}{N} \cdot \frac{N}{2} \right\} \\ &= \frac{\sqrt{\log p}}{\sqrt{N}} \cdot e^{y/2} \rightarrow \sqrt{\beta} e^{y/2} \end{aligned}$$

as  $n \rightarrow \infty$ . Collect all the facts above to have

$$\lim_{n \rightarrow \infty} h_n = K(\beta) e^{y/2}$$

where

$$K(\beta) = \left( \frac{\beta}{2\pi(1 - e^{-4\beta})} \right)^{1/2}.$$

By (46) then Lemma 6.4 we have

$$\lim_{n \rightarrow \infty} P((n-2) \log(1 - L_n^2) + 4 \log p - \log \log p \geq y) = e^{-K(\beta) e^{y/2}}$$

for any  $y \in \mathbb{R}$ . By the same argument as getting (39), the above yields that

$$\lim_{n \rightarrow \infty} P((n-2) \log(1 - L_n^2) + 4 \log p - \log \log p \leq y) = 1 - e^{-K(\beta) e^{y/2}}$$

for any  $y \in \mathbb{R}$ . We eventually arrive at (45). ■

**Proof of Theorem 3.** (i). Assuming (ii) of the theorem, dividing (13) by  $\log p$ , we see that

$$\frac{n}{\log p} \log(1 - L_n^2) \rightarrow -4$$

in probability as  $n \rightarrow \infty$ . Since  $(\log p)/n \rightarrow 0$ , we have  $L_n \rightarrow 1$  in probability as  $n \rightarrow \infty$ .

(ii). The proof in this part does not rely on the conclusion in (i). Fix  $y \in \mathbb{R}$ . Let  $N = n - 2$  and  $t = t_n \geq 0$  such that

$$t^2 = 1 - \exp \left\{ \frac{1}{N} (-4 \log p + \log n + y) \wedge 0 \right\}.$$

Obviously,  $t_n \rightarrow 1^-$  as  $n \rightarrow \infty$  by the condition  $(\log p)/n \rightarrow 0$ . Thus, without loss of generality, assume  $t = t_n \in (0, 1)$  for all  $n \geq 1$ . Easily,

$$\log(1 - t^2) = \frac{-4 \log p + \log n + y}{N} \quad \text{and} \quad (49)$$

$$P((n-2) \log(1 - L_n^2) + 4 \log p - \log n \geq y) = P(L_n \leq t) \quad (50)$$

as  $n$  is sufficiently large. We now evaluate  $h_n$  in Lemma 6.4 to obtain  $\lim_{n \rightarrow \infty} P(L_n \leq t)$ . Recall

$$h_n = \frac{n^{1/2} p^2}{\sqrt{2\pi}} \int_t^1 (1 - x^2)^{\frac{n-4}{2}} dx. \quad (51)$$

From Lemma 6.2 and the fact  $t_n \rightarrow 1$  as  $n \rightarrow \infty$  we obtain

$$\begin{aligned} n^{1/2} p^2 \int_t^1 (1-x^2)^{(n-4)/2} dx &\sim \frac{n^{1/2} p^2}{nt} (1-t^2)^{(n-2)/2} \\ &\sim \frac{p^2}{\sqrt{N}} (1-t^2)^{N/2} \end{aligned}$$

as  $n \rightarrow \infty$ . Combine this and (49) to have

$$\begin{aligned} \frac{p^2}{\sqrt{N}} (1-t^2)^{N/2} &\sim \frac{p^2}{\sqrt{N}} \cdot \exp \left\{ \frac{-4 \log p + \log n + y}{N} \cdot \frac{N}{2} \right\} \\ &= e^{y/2} \cdot \frac{\sqrt{n}}{\sqrt{N}} \rightarrow e^{y/2} \end{aligned}$$

as  $n \rightarrow \infty$ . Joining all the above we have that

$$\lim_{n \rightarrow \infty} h_n = \frac{1}{\sqrt{2\pi}} e^{y/2}$$

as  $n \rightarrow \infty$ . From (50) then Lemma 6.4 we finally obtain

$$\lim_{n \rightarrow \infty} P((n-2) \log(1-L_n^2) + 4 \log p - \log n \geq y) = e^{-K e^{y/2}}$$

for any  $y \in \mathbb{R}$ , where  $K = \frac{1}{\sqrt{2\pi}}$ . By the same argument as getting (39), the above actually implies that

$$\lim_{n \rightarrow \infty} P((n-2) \log(1-L_n^2) + 4 \log p - \log n \leq y) = 1 - e^{-K e^{y/2}}$$

for any  $y \in \mathbb{R}$ . This says that

$$(n-2)T_n + 4 \log p - \log n \implies F(y) \tag{52}$$

with  $F(y) = 1 - e^{-K e^{y/2}}$ ,  $y \in \mathbb{R}$  and  $K = 1/\sqrt{2\pi}$ . Further, multiplying the left hand side of (52) by  $\frac{2}{n-2}$  we obtain

$$2T_n + \frac{8 \log p}{n-2} \xrightarrow{P} 0 \tag{53}$$

as  $n \rightarrow \infty$ . Noticing  $(n-2)T_n + 2T_n = nT_n$ . Adding up (52) and (53), we conclude from the Slutsky lemma that

$$nT_n + 4 \log p - \log n + \frac{8 \log p}{n-2} = nT_n + \frac{4n}{n-2} \log p - \log n$$

converges weakly to the distribution function  $F(y) = 1 - e^{-K e^{y/2}}$ ,  $y \in \mathbb{R}$  with  $K = 1/\sqrt{2\pi}$ .

■



### 6.3 Proofs for Results on $\tilde{L}_n$ in Section 2.2

The proofs of the results on  $\tilde{L}_n$  are analogous to those of the results on  $L_n$ . The essential difference is to apply (22) in place of (21). Keeping all other arguments, we then get the proofs of the results on  $\tilde{L}_n$  stated in Section 2.2. We omit the details for reasons of space.

## References

- [1] Ahlfors, L. V. (1979). *Complex Analysis*, Third Edition. McGraw-Hill, New York.
- [2] Arratia, R. and Goldstein, L. and Gordon, L. (1989). Two moments suffice for Poisson approximation: The Chen-Stein method. *Ann. Probab.* **17**, 9-25.
- [3] Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37**, 1705-1732.
- [4] Cai, T. T. and Jiang, T. (2010). Limiting Laws of Coherence of Random Matrices with Applications to Testing Covariance Structure and Construction of Compressed Sensing Matrices. *Ann. Statist.*, to appear.
- [5] Cai, T. T. and Lv, J. (2007). Discussion: The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35**, 2365-2369.
- [6] Cai, T. T., Wang, L. and Xu, G. (2010a). Shifting inequality and recovery of sparse signals, *IEEE Transactions on Signal Processing* **58**, 1300-1308.
- [7] Cai, T. T., Wang, L. and Xu, G. (2010b). Stable recovery of sparse signals and an oracle inequality, *IEEE Transactions on Information Theory* **56**, 3516-3522.
- [8] Candès, E. J. and Plan, Y. (2009). Near-ideal model selection by  $\ell_1$  minimization. *Ann. Statist.* **37**, 2145-2177.
- [9] Candès, E. J. and Tao, T. (2007). The Dantzig selector: statistical estimation when  $p$  is much larger than  $n$  (with discussion). *Ann. Statist.* **35**, 2313-2351.
- [10] Candès, E. J. and Tao, T. (2005). Decoding by linear programming, *IEEE Trans. Inf. Theory* **51**, 4203-4215.
- [11] Donoho, D. (2006). Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289-1306.
- [12] Donoho, D. L., Elad, M. and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **52**, 6-18.

- [13] Donoho, D. L. and Huo, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory* **47**, 2845-2862.
- [14] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. Roy. Statist. Soc. Ser. B* 70, 849-911.
- [15] Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* 20, 101-148.
- [16] Fuchs, J. J. (2004). On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory* **50**, 1341-1344.
- [17] Gamelin, T. W. (2001). *Complex Analysis*. Springer.
- [18] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14**, 865-880.
- [19] Kariya, T. and Eaton, M. L. (1977). Robust tests for spherical symmetry. *Ann. Statist.* **5**, 206-215.
- [20] Li, D., Liu, W. and Rosalsky, A. (2009). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Relat. Fields* **148**, 5-35.
- [21] Li, D., Qi, Y. and Rosalsky, A. (2010). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. Preprint.
- [22] Li, D. and Rosalsky, A. (2006). Some strong limit theorems for the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **16**, 423-447.
- [23] Liu, W., Lin, Z. and Shao, Q. (2008). The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18**, 2337-2366.
- [24] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- [25] Wigner, E. P. (1958). On the distribution of roots of certain symmetric matrices. *Ann. Math.* **67**, 325-328.
- [26] Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transaction of American Mathematical Society* **359**, 5345-5363.